

ONE-BIT COMPRESSED SENSING WITH NON-GAUSSIAN MEASUREMENTS

ALBERT AI, ALEX LAPANOWSKI, YANIV PLAN, AND ROMAN VERSHYNIN

ABSTRACT. In one-bit compressed sensing, previous results state that sparse signals may be robustly recovered when the measurements are taken using Gaussian random vectors. In contrast to standard compressed sensing, these results are not extendable to natural non-Gaussian distributions without further assumptions, as can be demonstrated by simple counter-examples. We show that approximately sparse signals, which also satisfy a mild infinity-norm constraint, can be accurately reconstructed from single-bit measurements sampled according to a sub-gaussian distribution, and the reconstruction comes as the solution to a convex program.

Keywords: 1-bit compressed sensing; quantization; signal reconstruction; convex programming

1. INTRODUCTION

In the standard noiseless compressed sensing model, one has access to linear measurements of the form

$$y_i = \langle \mathbf{a}_i, \mathbf{x} \rangle, \quad i = 1, 2, \dots, m$$

where $\mathbf{a}_1, \dots, \mathbf{a}_m \in \mathbb{R}^n$ are known measurement vectors and $\mathbf{x} \in \mathbb{R}^n$ is a sparse signal which one wishes to reconstruct (see e.g. [2]). Let $\|\mathbf{x}\|_0$ denote the number of nonzero entries in \mathbf{x} . Typical results state that when the measurement vectors are chosen randomly from a sub-gaussian distribution, and $\|\mathbf{x}\|_0 \leq s$, then $m = O(s \log(n/s))$ measurements are sufficient for robust recovery of the signal \mathbf{x} (see [2]).

In one-bit compressed sensing, the measurements are compressed to single bits, and thus they take the form

$$(1.1) \quad y_i = \text{sign}(\langle \mathbf{a}_i, \mathbf{x} \rangle), \quad i = 1, 2, \dots, m.$$

Here, the sign function is defined by $\text{sign}(t) = 1$ when $t \geq 0$ and -1 otherwise. Clearly, the magnitude of \mathbf{x} is lost in these measurements and so the goal is to approximate the direction of \mathbf{x} . Thus we may assume without loss of generality that $\mathbf{x} \in S^{n-1}$.

One-bit compressed sensing was introduced in [1] to model extreme quantization in compressed sensing; the webpage <http://dsp.rice.edu/1bitCS/> details its practical applications and the recent literature. We also note the similarity in model to sparse logistic regression; the connection

Date: Submitted August 2012.

2000 *Mathematics Subject Classification.* 94A12; 60D05; 90C25.

Y.P. was supported by an NSF Postdoctoral Research Fellowship under award No. 1103909. R.V. was supported by NSF grant 1001829. A.A. and A.L. were REU students supported by the NSF grant 0918623.

will be made clear in Section 1.1. We review the previous results here; we note that there are algorithmic results, theoretical results, and results which consider quantization with more than two bits (see the above webpage for more details). We restrict our review to the theoretical results that consider one-bit quantization.

Suppose that the signal $\mathbf{x} \in \mathbb{R}^n$ satisfies $\|\mathbf{x}\|_0 \leq s$. Gupta et al. [3] assume that the measurement vectors are Gaussian and demonstrate that the support of \mathbf{x} can tractably be recovered from either 1) $O(s \log n)$ nonadaptive measurements assuming a constant dynamic range of \mathbf{x} (i.e. the magnitude of all nonzero entries of \mathbf{x} is assumed to lie between two constants), or 2) $O(s \log n)$ adaptive measurements. Jacques et al. [4] introduce a certain *binary ϵ -stable embedding property* which is a one-bit analogue to the *restricted isometry property* of standard compressed sensing. They demonstrate that Gaussian measurement ensembles satisfy this property with high probability (given enough measurements). Assuming the binary ϵ -stable embedding property holds, they show that any estimate of \mathbf{x} which is both s -sparse and approximately matches the data, will be accurate. In particular, $O(s \log n)$ Gaussian measurements are sufficient to have a relative error bounded by any fixed constant. These results are robust to noise.

Plan and Vershynin [7, 8] show that one may reconstruct a sparse signal \mathbf{x} from single-bit measurements by *convex programming*, for which tractable solvers exist. [7] considers the noiseless case and [8] considers the noisy case (and also sparse logistic regression). In [8] and the present paper, the model for the signal \mathbf{x} is allowed to be quite general, with sparsity as a special case. Indeed, suppose \mathbf{x} belongs to some known set K , which is meant to encode the *model of the signal structure*. For example, in order to encode sparsity, one could let K be the set

$$S_{n,s} := \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_0 \leq s, \|\mathbf{x}\|_2 \leq 1\}.$$

The recovery is achieved in [8] by solving the optimization problem

$$(1.2) \quad \max \sum_{i=1}^m y_i \langle \mathbf{a}_i, \mathbf{x}' \rangle \quad \text{subject to} \quad \mathbf{x}' \in K.$$

If K is a convex set then (1.2) is a convex optimization problem, so it can be solved by a variety of convex optimization solvers.

However, the reader may note that the set of sparse vectors $S_{n,s}$ is extremely non-convex. To overcome this, it was proposed in [8] to take K to be an approximate convex relaxation of $S_{n,s}$ (see [7, Lemma 3.1]), namely

$$(1.3) \quad K = K_{n,s} := \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_2 \leq 1, \|\mathbf{x}\|_1 \leq \sqrt{s}\}.$$

It was shown in [8] that $m = O(s \log(n/s))$ Gaussian measurements are sufficient to accurately recover \mathbf{x} by solving the convex optimization problem (1.2).

A natural question is whether reconstruction of \mathbf{x} from one-bit measurements is still feasible when measurements are taken using random vectors with *non-Gaussian* coordinates. A simple counterexample shows that this is not generally possible even when the coordinates are sub-gaussian. Suppose

that all coordinates of \mathbf{a}_i are in $\{-1, 1\}$. For example, one may let the coordinates be independent symmetric Bernoulli random variables. Then the vectors

$$\mathbf{x} = (1, \frac{1}{2}, 0, \dots, 0) \quad \text{and} \quad \mathbf{x}' = (1, -\frac{1}{2}, 0, \dots, 0)$$

clearly satisfy $\text{sign}(\langle \mathbf{a}_i, \mathbf{x} \rangle) = \text{sign}(\langle \mathbf{a}_i, \mathbf{x}' \rangle) = 1$. This shows that one can not distinguish the two very different signals \mathbf{x} and \mathbf{x}' by such measurements,¹ even if infinitely many measurements are taken.

One may ask whether this counterexample is typical or worst-case behavior. In this paper, we demonstrate that the latter is the case—a *difficulty can only arise for extremely sparse signals*. Namely, we show that under the assumption

$$(1.4) \quad \|\mathbf{x}\|_\infty \ll \|\mathbf{x}\|_2 = 1,$$

an approximate recovery of \mathbf{x} is still possible with general sub-gaussian measurements, and it is achieved by the convex program (1.2). Furthermore, we prove that for the distributions that are near Gaussian (in total variation), an approximate recovery of \mathbf{x} is possible even without the assumption (1.4).

1.1. Main Results. We shall assume that the signal set K lies in the unit Euclidean ball in \mathbb{R}^n , which we shall denote B_2^n . The quality of recovery of a signal $\mathbf{x} \in K$ will depend on K through a single geometric parameter – the *Gaussian mean width* of K . It is defined as

$$w(K) = \mathbb{E} \sup_{\mathbf{x} \in K-K} \langle \mathbf{g}, \mathbf{x} \rangle,$$

where \mathbf{g} denotes a standard Gaussian random vector in \mathbb{R}^n , i.e. a vector with independent $N(0, 1)$ random coordinates. The reader may refer to [8, Section 2] for a brief overview of the properties of mean width.

The main purpose of this paper is to allow the measurement vectors \mathbf{a}_i to have general *sub-gaussian* (rather than Gaussian) independent coordinates. Recall that a random variable a is sub-gaussian if its distribution is dominated by a centered normal distribution. This property can be expressed in several equivalent ways, see [11, Section 5.2.3]. One convenient way to define a sub-gaussian random variable is to require that its moments be bounded by the corresponding moments of $N(0, 1)$, so that $(\mathbb{E} |a|^p)^{1/p} = O(\sqrt{p})$ as $p \rightarrow \infty$. Formally, a is called sub-gaussian if

$$(1.5) \quad \kappa := \sup_{p \geq 1} p^{-1/2} (\mathbb{E} |a|^p)^{1/p} < \infty.$$

The quantity κ is called the *sub-gaussian norm* of a . The class of sub-gaussian random variables includes in particular normal, Bernoulli and all bounded random variables.

¹One can normalize the signals \mathbf{x} and \mathbf{x}' to lie on S^1 , and the same phenomenon clearly persists.

Our main result is a generalization of [8, Theorem 1.1], which states that when the measurement vectors $\mathbf{a}_1, \dots, \mathbf{a}_m$ are Gaussian, then

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 \lesssim \frac{w(K)}{\sqrt{m}}$$

with high probability. Our following result generalizes this theorem to when the measurement vectors $\mathbf{a}_1, \dots, \mathbf{a}_m$ have coordinates with sub-gaussian distribution. The only important difference is that the error now has an additive dependence on $\|\mathbf{x}\|_\infty$. This serves to exclude extremely sparse signals, which can destroy recovery, according to the example we discussed above.

Theorem 1.1 (Estimating a signal with no noise). *Let $a \in \mathbb{R}$ be a symmetric, sub-gaussian, and unit variance random variable with κ as in (1.5). Let $\mathbf{a}_1, \dots, \mathbf{a}_m$ be independent random vectors in \mathbb{R}^n whose coordinates are i.i.d. copies of a . Consider signal set $K \subseteq B_2^n$, and fix $\mathbf{x} \in K$ satisfying $\|\mathbf{x}\|_2 = 1$. Let \mathbf{y} follow the 1-bit measurement model of Equation (1.1). Then for each $\beta > 0$, with probability at least $1 - 4e^{-\beta^2}$, the solution $\hat{\mathbf{x}}$ to the optimization problem (1.2) satisfies*

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 \leq C \left(\kappa^3 \|\mathbf{x}\|_\infty^{1/2} + \frac{\kappa}{\sqrt{m}} (w(K) + \beta) \right).$$

In this theorem and later, C and c denote positive absolute constants, which can be different from line to line.

A proof of Theorem 1.1 is given in Section 3.

This theorem can be easily specialized to sparse (and approximately sparse) signals. To this end, we consider $K = K_{n,s}$ as in (1.3). A standard computation (see [8, Equation 3.3]) shows that

$$w(K_{n,s}) \leq C \sqrt{s \log(2n/s)}.$$

Then the following corollary follows directly from Theorem 1.1.

Corollary 1.2 (Estimating a sparse signal with no noise). *Let $K = K_{n,s}$, $s \geq 1$, and let everything else be as in Theorem 1.1. Then with probability at least $1 - 4 \exp \{-2s \log(2n/s)\} \geq 1 - \frac{1}{n^2}$, the solution $\hat{\mathbf{x}}$ to the optimization problem (1.2) satisfies*

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 \leq C \left(\kappa^3 \|\mathbf{x}\|_\infty^{1/2} + \kappa \sqrt{\frac{s \log(n/s)}{m}} \right).$$

In words, this result yields that if the signal is approximately s -sparse, but not extremely sparse so that $\|\mathbf{x}\|_\infty \ll \|\mathbf{x}\|_2 = 1$, then with high probability \mathbf{x} can be accurately recovered from

$$m = O(s \log(n/s))$$

general sub-gaussian measurements.

We also establish a version of Theorem 1.1 under a statistical or noisy model. A noisy random measurement is modeled by a random variable y_i taking values in $\{-1, 1\}$ such that

$$(1.6) \quad \mathbb{E}(y_i | \mathbf{a}_i) = \theta(\langle \mathbf{a}_i, \mathbf{x} \rangle), \quad i = 1, 2, \dots, m$$

where $\theta(\cdot)$ is some function, which may even be unknown or unspecified. We only assume that $\theta(t) \in C^3(\mathbb{R})$, the first three derivatives being bounded by τ_1, τ_2, τ_3 respectively, and that

$$(1.7) \quad \mathbb{E} \theta(g)g =: \lambda > 0$$

where $g \propto N(0, 1)$. For example, in sparse logistic regression one would take

$$\theta(t) = \tanh(t/2),$$

with bounds $\tau_1 = 0.5$, $\tau_2 \approx 0.19$, $\tau_3 \approx 0.083$ and $\lambda \approx 0.41$.

To note another important example, observe that the setting of Theorem 1.1 is described by choosing $\theta(t) = \text{sign}(t)$ and disregarding the differentiability requirements. In this case, $\lambda = \sqrt{2/\pi}$. It is useful to note that this is the largest possible value λ can take, over all possible $\theta(t)$.

The following is a version of Theorem 1.1 under this noisy or statistical model.

Theorem 1.3 (Estimating a spread signal under random noise). *We remain in the setting of Theorem 1.1, but with random measurements y_i modeled as in Equation (1.6). Then for each $\beta > 0$, with probability at least $1 - 4e^{-\beta^2}$, the solution $\hat{\mathbf{x}}$ to the optimization problem (1.2) satisfies*

$$(1.8) \quad \|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 \leq C \left(\frac{\kappa^2}{\lambda^{3/2}} ((\tau_2 + \tau_3)(\tau_1 + 1) \|\mathbf{x}\|_\infty)^{1/2} + \frac{\kappa}{\lambda\sqrt{m}} (w(K) + \beta) \right).$$

For Gaussian measurement vectors \mathbf{a}_i , a version of this theorem was proved in [8].

The proof of Theorem 1.3 is provided in Section 2.

An interested reader may specialize this result to sparse signals \mathbf{x} as we did before, i.e. by taking $K = K_{n,s}$ and noting as in Corollary 1.2 that $w(K_{n,s}) \leq C\sqrt{\log(2n/s)}$.

Our last result is about non-Gaussian distributions, which nevertheless are close to Gaussian in total variation. For such measurements, it is reasonable to expect that the same conclusions as for Gaussian measurements, i.e. that the theorems above hold for all signals \mathbf{x} without any dependence on $\|\mathbf{x}\|_\infty$. We confirm that this is the case. Suppose that the coordinates of \mathbf{a}_i are i.i.d. copies of a random variable a that satisfies the total variation bound

$$\|a - g\|_{TV} := \sup_A |P(a \in A) - P(g \in A)| \leq \varepsilon$$

where $g \propto N(0, 1)$. In the case when $\theta(t) = \text{sign}(t)$, one has

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 \lesssim \varepsilon^{1/8} + \frac{w(K)}{\sqrt{m}},$$

and in the case when $\theta(t) \in C^2$ one has

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 \lesssim \varepsilon^{1/2} + \frac{w(K)}{\sqrt{m}}.$$

The precise results and their proofs are provided in the appendix as Theorems 4.1 and 4.4 respectively.

2. PROOF OF THEOREM 1.3

It will be convenient to define the (rescaled) objective function for our convex program (1.2):

$$f_{\mathbf{x}}(\mathbf{x}') := \frac{1}{m} \sum_{i=1}^m y_i \langle \mathbf{a}_i, \mathbf{x}' \rangle.$$

We reduce the proof to two main propositions.

Proposition 2.1 (Expectation). *Consider $\mathbf{x} \in S^{n-1}$, $\mathbf{x}' \in B_2^n$. If \mathbf{x} satisfies*

$$(2.1) \quad \|\mathbf{x}\|_{\infty} \leq \frac{\lambda}{C(\tau_2 + \tau_3) \mathbb{E} a^4},$$

then

$$(2.2) \quad |\mathbb{E} f_{\mathbf{x}}(\mathbf{x}') - \langle \lambda \mathbf{x}, \mathbf{x}' \rangle| \leq \frac{C'}{\lambda^{1/2}} ((\tau_2 + \tau_3)(\tau_1 + 1) \mathbb{E} a^4 \|\mathbf{x}\|_{\infty})^{1/2}.$$

The proof of Proposition 2.1 is provided in Section 2.1.

Proposition 2.2 (Concentration). *For each $t > 0$,*

$$P \left(\sup_{\mathbf{z} \in K-K} |f_{\mathbf{x}}(\mathbf{z}) - \mathbb{E} f_{\mathbf{x}}(\mathbf{z})| \geq C\kappa \frac{w(K) + \beta}{\sqrt{m}} \right) \leq 4e^{-\beta^2}.$$

The proof of Proposition 2.2 is provided in Section 2.2.

Proof of Theorem 1.3. First observe that if

$$\|\mathbf{x}\|_{\infty} > \frac{\lambda}{C(\tau_2 + \tau_3) \mathbb{E} a^4}$$

then one may show that the right-hand side of Equation (1.8) is greater than 4, and the theorem trivially follows. (In verifying this calculation, note that $\kappa^4 / \mathbb{E} a^4 \geq 1/16$ and $1/\lambda \geq \sqrt{\pi/2}$.)

Hence, we may suppose otherwise, in which case Proposition 2.1 applies. Consider $\mathbf{z}' = \mathbf{x}' - \mathbf{x} \in K - K$. Further, using Proposition 2.1, we find

$$\begin{aligned} -\mathbb{E} f_{\mathbf{x}}(\mathbf{z}') &= \mathbb{E} f_{\mathbf{x}}(\mathbf{x}) - \mathbb{E} f_{\mathbf{x}}(\mathbf{x}') \geq \langle \lambda \mathbf{x}, \mathbf{x} \rangle - \langle \lambda \mathbf{x}, \mathbf{x}' \rangle - 2 \frac{C}{\lambda^{1/2}} ((\tau_2 + \tau_3)(\tau_1 + 1) \mathbb{E} a^4 \|\mathbf{x}\|_{\infty})^{1/2} \\ &\geq \frac{\lambda}{2} \|\mathbf{x} - \mathbf{x}'\|_2^2 - \frac{C}{\lambda^{1/2}} ((\tau_2 + \tau_3)(\tau_1 + 1) \mathbb{E} a^4 \|\mathbf{x}\|_{\infty})^{1/2}. \end{aligned}$$

By Proposition 2.2, we have the following event and a lower bound on its probability, respectively:

$$\sup_{\mathbf{z} \in K-K} |f_{\mathbf{x}}(\mathbf{z}) - \mathbb{E} f_{\mathbf{x}}(\mathbf{z})| \leq C\kappa \frac{w(K) + \beta}{\sqrt{m}}, \quad 1 - 4e^{-\beta^2}.$$

In this event, note that

$$\begin{aligned} f_{\mathbf{x}}(\mathbf{z}') &\leq \mathbb{E} f_{\mathbf{x}}(\mathbf{z}') + C\kappa \frac{w(K) + \beta}{\sqrt{m}} \\ &\leq \frac{C}{\lambda^{1/2}} ((\tau_2 + \tau_3)(\tau_1 + 1) \mathbb{E} a^4 \|\mathbf{x}\|_{\infty})^{1/2} - \frac{\lambda}{2} \|\mathbf{x} - \widehat{\mathbf{x}}\|_2^2 + C\kappa \frac{w(K) + \beta}{\sqrt{m}}. \end{aligned}$$

This holds uniformly for all $\mathbf{x}' \in K$. Pick $\mathbf{x}' = \hat{\mathbf{x}}$ and recall that $\hat{\mathbf{x}}$ maximizes $f_{\mathbf{x}}$; thus $f_{\mathbf{x}}(\mathbf{z}') = f_{\mathbf{x}}(\hat{\mathbf{x}}) - f_{\mathbf{x}}(\mathbf{x}) \geq 0$. Thus the right-hand side of the above inequality is bounded below by 0. By definition of κ , $\mathbb{E} a^4$ is bounded by $16\kappa^4$. Rearranging completes the proof of the theorem. ■

2.1. Expectation: Proof of Proposition 2.1. For convenience, let us denote $y := y_1$ and $\mathbf{a} := \mathbf{a}_1$. Recalling (1.6), we observe the following equivalences:

$$(2.3) \quad \mathbb{E} f_{\mathbf{x}}(\mathbf{x}') = \frac{1}{m} \sum_{i=1}^m \mathbb{E} y_i \langle \mathbf{a}_i, \mathbf{x}' \rangle = \mathbb{E} y \langle \mathbf{a}, \mathbf{x}' \rangle = \mathbb{E} (\mathbb{E} y \langle \mathbf{a}, \mathbf{x}' \rangle | \mathbf{a}) = \mathbb{E} \theta(\langle \mathbf{a}, \mathbf{x} \rangle) \langle \mathbf{a}, \mathbf{x}' \rangle.$$

In order to analyze the above quantity, we will compare to the case when \mathbf{a} is Gaussian, in which case the analysis is fairly simple, see [8, Section 4.1]. Such a comparison is a bi-variate version of Bery-Esseen central limit theorem for the function $\theta(\langle \mathbf{a}, \mathbf{x} \rangle) \langle \mathbf{a}, \mathbf{x}' \rangle$.

Lemma 2.3 (Berry-Esseen type central limit theorem). *Consider $\mathbf{x}, \mathbf{z} \in S^{n-1}$. Let $|\mathbf{x}|, |\mathbf{z}|$ be the vectors obtained by taking absolute values of the coordinates of \mathbf{x}, \mathbf{z} respectively. Then*

$$|\mathbb{E} \theta(\langle \mathbf{a}, \mathbf{x} \rangle) \langle \mathbf{a}, \mathbf{z} \rangle - \mathbb{E} \theta(\langle \mathbf{g}, \mathbf{x} \rangle) \langle \mathbf{g}, \mathbf{z} \rangle| \leq C(\tau_2 + \tau_3) \mathbb{E} a^4 \| |\mathbf{x}| + |\mathbf{z}| \|_3^3.$$

The proof is based on a Lindeberg replacement argument in two variables; it is provided in the appendix.

A challenge arises when we wish to apply Lemma 2.3 to the expectation $\mathbb{E} \theta(\langle \mathbf{a}, \mathbf{x} \rangle) \langle \mathbf{a}, \mathbf{x}' \rangle$ in (2.3). Indeed, we have no way to control $\|\mathbf{x}'\|_3$, the quantity that is crucial in bounding the difference in Lemma 2.3.² Recall that we have to treat all vectors $\mathbf{x}' \in K \subseteq B_2^n$ that arise in the optimization problem (1.2). Some of these vectors may be very sparse, having $\|\mathbf{x}'\|_3 \approx 1$, which produces a useless bound in Lemma 2.3.

Nevertheless, this obstacle can be bypassed. Observe that both sides of identity (2.3) are linear in \mathbf{x}' . This motivates us to define the vector

$$\mathbf{v}_x := \mathbb{E} \theta(\langle \mathbf{a}, \mathbf{x} \rangle) \cdot \mathbf{a}$$

and express

$$\mathbb{E} f_{\mathbf{x}}(\mathbf{x}') = \langle \mathbf{v}_x, \mathbf{x}' \rangle.$$

The conclusion of Proposition 2.1 states that

$$(2.4) \quad \mathbf{v}_x \approx \lambda \mathbf{x},$$

with the error bound given by the right hand side of (2.2).

This vector approximation may be difficult to prove directly, i.e. based on Bery-Esseen type central limit theorem in n dimensions. However, (2.4) clearly follows from the two *scalar* approximate identities:

$$(2.5) \quad \langle \mathbf{v}_x, \mathbf{x} \rangle \approx \lambda \quad \text{and} \quad \|\mathbf{v}_x\|_2 \approx \lambda.$$

²Another problem is the required normalization $\mathbf{x}' \in S^{n-1}$, but it is just a minor nuisance which can be addressed by rescaling.

(Indeed, the first of these approximate identities states that \mathbf{v}_x is near a hyperplane with normal \mathbf{x} , and the second one states that \mathbf{v}_x is near the centered sphere tangent to that hyperplane.) We reduced the problem to proving (2.5).

We shall deduce both inequalities in (2.5) from the Berry-Esseen Lemma 2.3 – the first inequality by choosing $\mathbf{z} = \mathbf{x}$, and the second by choosing $\mathbf{z} = \mathbf{v}_x / \|\mathbf{v}_x\|_2$. The first inequality is simple.

Lemma 2.4.

$$|\langle \mathbf{v}_x, \mathbf{x} \rangle - \lambda| \leq C(\tau_2 + \tau_3) \mathbb{E} a^4 \|\mathbf{x}\|_3^3.$$

Proof. By definition of \mathbf{x} , we have $\langle \mathbf{v}_x, \mathbf{x} \rangle = \mathbb{E} \theta(\langle \mathbf{a}, \mathbf{x} \rangle) \langle \mathbf{a}, \mathbf{x} \rangle$. On the other hand, by rotation invariance of Gaussian distribution and by condition (1.7), $\mathbb{E} \theta(\langle \mathbf{g}, \mathbf{x} \rangle) \langle \mathbf{g}, \mathbf{x} \rangle = \lambda$. Thus, the lemma is a special case of Lemma 2.3 when $\mathbf{z} = \mathbf{x}$. ■

To apply Berry-Esseen Lemma 2.3 for $\mathbf{z} = \mathbf{v}_x / \|\mathbf{v}_x\|_2$ in order to prove the second inequality in (2.5), we need to know a two-sided bound on $\|\mathbf{v}_x\|_2$ and an upper bound on $\|\mathbf{v}_x\|_\infty$. We establish them in the following two lemmas.

Lemma 2.5. $\lambda/2 \leq \|\mathbf{v}_x\|_2 \leq 1$.

Proof. For the lower bound, using Lemma 2.4, we have

$$(2.6) \quad \|\mathbf{v}_x\|_2 = \|\mathbf{v}_x\|_2 \|\mathbf{x}\|_2 \geq |\langle \mathbf{v}_x, \mathbf{x} \rangle| \geq \lambda - C(\tau_2 + \tau_3) \mathbb{E} a^4 \|\mathbf{x}\|_3^3.$$

Suppose the condition (2.1) of Proposition 2.1 holds with this C , then

$$\|\mathbf{x}\|_3^3 \leq \|\mathbf{x}\|_2^2 \|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_\infty \leq \frac{\lambda}{2C(\tau_2 + \tau_3) \mathbb{E} a^4}.$$

Plugging this into (2.6), we obtain $\|\mathbf{v}_x\|_2 \geq \lambda/2$, as desired.

In the other direction, we have

$$\|\mathbf{v}_x\|_2^2 = \langle \mathbf{v}_x, \mathbf{v}_x \rangle = \mathbb{E} \theta(\langle \mathbf{a}, \mathbf{x} \rangle) \langle \mathbf{a}, \mathbf{v}_x \rangle.$$

Recall from the definition (1.6) that the function θ is automatically bounded by 1, so

$$\|\mathbf{v}_x\|_2^2 \leq \mathbb{E} |\langle \mathbf{a}, \mathbf{v}_x \rangle| \leq (\mathbb{E} \langle \mathbf{a}, \mathbf{v}_x \rangle^2)^{1/2} = (\|\mathbf{v}_x\|_2^2 \mathbb{E} a_1^2)^{1/2} = \|\mathbf{v}_x\|_2.$$

It follows that $\|\mathbf{v}_x\|_2 \leq 1$, as desired. ■

Lemma 2.6. $\|\mathbf{v}_x\|_\infty \leq \tau_1 \|\mathbf{x}\|_\infty$.

Proof. We have $\|\mathbf{v}_x\|_\infty = \max_{i \in [n]} |\mathbb{E} \theta(\langle \mathbf{a}, \mathbf{x} \rangle) a_i|$. Let us express in coordinates

$$\langle \mathbf{a}, \mathbf{x} \rangle = \sum_{k=1}^n a_k x_k = S + a_i x_i$$

where we denote $S = \sum_{k \neq i} a_k x_k$. By assumption, the distribution of a_i is symmetric. Therefore a_i is identically distributed with $|a_i|\varepsilon$, where ε denotes an independent symmetric Bernoulli random variable. Conditioning on a_i , we obtain

$$\mathbb{E} \theta(\langle \mathbf{a}, \mathbf{x} \rangle) a_i = \mathbb{E} \theta(S + |a_i| \varepsilon x_i) |a_i| \varepsilon = \frac{1}{2} \mathbb{E} [\theta(S + |a_i| x_i) - \theta(S - |a_i| x_i)] |a_i|.$$

By the Mean Value Theorem, $\frac{1}{2} |\theta(S + t) - \theta(S - t)| \leq \|\theta'\|_\infty t \leq \tau_1 t$ for $t > 0$. Thus

$$|\mathbb{E} \theta(\langle \mathbf{a}, \mathbf{x} \rangle) a_i| \leq \tau_1 |x_i| \mathbb{E} |a_i|^2 = \tau_1 |x_i|.$$

The conclusion of the lemma easily follows. ■

We can now turn to the second approximate identity in (2.5). We will only need an upper bound, and we formally state it in the following lemma.

Lemma 2.7.

$$\|\mathbf{v}_x\|_2 \leq \lambda + \frac{C}{\lambda} (\tau_2 + \tau_3) (\tau_1 + \lambda) \mathbb{E} a^4 \|\mathbf{x}\|_\infty.$$

Proof. We express

$$\|\mathbf{v}_x\|_2 = \langle \mathbf{v}_x, \mathbf{v}_x / \|\mathbf{v}_x\|_2 \rangle = \mathbb{E} \theta(\langle \mathbf{a}, \mathbf{x} \rangle) \langle \mathbf{a}, \mathbf{z} \rangle, \quad \text{where } \mathbf{z} := \mathbf{v}_x / \|\mathbf{v}_x\|_2.$$

We would like to apply the Berry-Esseen type result, Lemma 2.3. The corresponding quantity for the normal distribution can be easily computed using rotation invariance, see [8, Lemma 4.1]:

$$\mathbb{E} \theta(\langle \mathbf{g}, \mathbf{x} \rangle) \langle \mathbf{g}, \mathbf{z} \rangle = \lambda \langle \mathbf{x}, \mathbf{z} \rangle.$$

Lemma 2.3 then yields

$$(2.7) \quad \|\mathbf{v}_x\|_2 \leq \lambda \langle \mathbf{x}, \mathbf{z} \rangle + C(\tau_2 + \tau_3) \mathbb{E} a^4 \|\mathbf{x}\| + \|\mathbf{z}\|_3^3.$$

Next, Lemma 2.5 and Lemma 2.6 together imply that $\|\mathbf{z}\|_\infty \leq 2\tau_1 \|\mathbf{x}\|_\infty / \lambda$. Hence

$$(2.8) \quad \begin{aligned} \frac{1}{4} \|\mathbf{x}\| + \|\mathbf{z}\|_3^3 &\leq \frac{1}{4} \|\mathbf{x}\| + \|\mathbf{z}\|_2^2 \cdot \|\mathbf{x}\| + \|\mathbf{z}\|_\infty \\ &\leq \|\mathbf{x}\| + \|\mathbf{z}\|_\infty = \|\mathbf{x}\|_\infty + \|\mathbf{z}\|_\infty \leq \frac{2\tau_1 + \lambda}{\lambda} \|\mathbf{x}\|_\infty. \end{aligned}$$

Combining (2.7) and (2.8), we complete the proof. ■

We are ready to deduce Proposition 2.1 from Lemmas 2.4 and 2.7.

Proof of Proposition 2.1.

$$|\mathbb{E} f_x(\mathbf{x}') - \langle \lambda \mathbf{x}, \mathbf{x}' \rangle| = |\langle \mathbf{v}_x, \mathbf{x}' \rangle - \langle \lambda \mathbf{x}, \mathbf{x}' \rangle| \leq \|\mathbf{v}_x - \lambda \mathbf{x}\|_2.$$

Next, expanding and rearranging the terms, we have

$$\|\mathbf{v}_x - \lambda \mathbf{x}\|_2^2 = \|\mathbf{v}_x\|_2^2 + \lambda^2 \|\mathbf{x}\|_2^2 - 2\lambda \langle \mathbf{v}_x, \mathbf{x} \rangle = (\|\mathbf{v}_x\|_2 + \lambda)(\|\mathbf{v}_x\|_2 - \lambda) + 2\lambda(\lambda - \langle \mathbf{v}_x, \mathbf{x} \rangle).$$

Now we bound all these terms. As we mentioned in the introduction, $\lambda \leq \sqrt{2/\pi}$, and by Lemma 2.5, $\|\mathbf{v}_x\|_2 \leq 1$. This controls the factor $\|\mathbf{v}_x\|_2 + \lambda$. Lemma 2.7 gives a bound on the factor $\|\mathbf{v}_x\|_2 - \lambda$. Finally, Lemma 2.4 gives a bound on $|\lambda - \langle \mathbf{v}_x, \mathbf{x} \rangle|$. Putting all these together, we conclude that

$$\|\mathbf{v}_x - \lambda \mathbf{x}\|_2^2 \leq C \left(\frac{1}{\lambda} (\tau_2 + \tau_3) (\tau_1 + \lambda) \mathbb{E} a^4 \|\mathbf{x}\|_\infty + (\tau_2 + \tau_3) \mathbb{E} a^4 \|\mathbf{x}\|_3^3 \right).$$

The last line follows from Lemmas 2.4 and 2.7. Since $\|\mathbf{x}\|_3^3 \leq \|\mathbf{x}\|_2^2 \|\mathbf{x}\|_\infty = \|\mathbf{x}\|_\infty$, this completes the proof. ■

2.2. Concentration: Proof of Proposition 2.2. We need to control the random variable

$$Z := \sup_{\mathbf{z} \in K-K} |f_{\mathbf{x}}(\mathbf{z}) - \mathbb{E} f_{\mathbf{x}}(\mathbf{z})|.$$

This will be done using techniques from probability in Banach spaces, following the argument in [8, Proposition 4.2]. The symmetrization lemma below allows us to essentially replace Z by the random variable

$$Z' := \sup_{\mathbf{z} \in K-K} \frac{1}{m} \left| \sum_{i=1}^m \varepsilon_i y_i \langle \mathbf{a}_i, \mathbf{z} \rangle \right|.$$

where ε_i denote independent symmetric Bernoulli random variables.

Lemma 2.8 (Symmetrization). *We have*

$$(2.9) \quad \mathbb{E} Z \leq 2 \mathbb{E} Z'.$$

Furthermore, for each $t > 0$ we have the deviation inequality

$$(2.10) \quad P(Z \geq 2 \mathbb{E} Z + t) \leq 4P(Z' > t/2).$$

The proof of this result is identical to the proof of [8, Lemma 5.1] for the normal distribution.

The following is a standard Gaussian concentration inequality, which is a simple extension of [5, Theorem 7.1].

Lemma 2.9 (Gaussian concentration). *Given a set $K \subseteq B_2^n$, we have*

$$P \left(\sup_{\mathbf{z} \in K-K} \langle \mathbf{g}, \mathbf{z} \rangle - w(K) > r \right) \leq e^{-r^2/8}, \quad r > 0.$$

The following inequality is an adaptation of [6, Lemma 4.6]:

Lemma 2.10 (Contraction Principle). *Consider sequences of independent symmetric random variables η_i and ξ_i such that for some scalar $M \geq 1$, and every i and $t > 0$,*

$$P(|\eta_i| > t) \leq MP(|\xi_i| > t).$$

Then for any finite sequence x_i and an integer $p \geq 1$, we have

$$\mathbb{E} \left(\left\| \sum_{i=1}^n \eta_i x_i \right\| \right)^p \leq \mathbb{E} \left(M \left\| \sum_{i=1}^n \xi_i x_i \right\| \right)^p.$$

We will first apply Lemma 2.10 to derive a moment bound on Z' , and then convert the moment bound back into a tail bound to apply in the right-hand side of Equation (2.10).

Because $\varepsilon_i y_i \mathbf{a}_i$ has the same distribution as \mathbf{a}_i , and by the symmetry of $K - K$,

$$\mathbb{E}(Z')^p = \mathbb{E} \left(\sup_{\mathbf{z} \in K-K} \frac{1}{m} \sum_{i=1}^m \langle \mathbf{a}_i, \mathbf{z} \rangle \right)^p = \mathbb{E} \left(\sup_{\mathbf{z} \in K-K} \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^n (a_i)_j z_j \right)^p.$$

We apply Lemma 2.10 with $(a_i)_j$ in place of η_i , $e_i e_j^*$ in place of x_i (where e_i is the i -th standard basis vector), ξ_i as independent $N(0, 1)$ random variables, and the matrix semi-norm defined by $\|A\| := \sup_{\mathbf{z} \in K-K} \sum_{i,j} A_{i,j} z_j$. To this end, recall that $(a_i)_j$ are distributed identically with a . Since a is a sub-gaussian random variable, it follows from definition (1.5) that

$$P(|a| > t) \leq CP(|g| \cdot \kappa > t), \quad t > 0.$$

Therefore an application of Lemma 2.10 allows us to replace $(a_i)_j$ by $(C\kappa)(g_i)_j$ and thus conclude that

$$(2.11) \quad \mathbb{E}(Z')^p \leq \mathbb{E} \left(\sup_{\mathbf{z} \in K-K} \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^n C\kappa(g_i)_j z_j \right)^p = \mathbb{E} \left(\frac{C\kappa}{\sqrt{m}} \sup_{\mathbf{z} \in K-K} \langle \mathbf{g}, \mathbf{z} \rangle \right)^p.$$

To further develop this inequality, we express the Gaussian concentration tail bound (Lemma 2.9) in terms of moment bounds. For convenience, define

$$\xi = \sup_{\mathbf{z} \in K-K} \langle \mathbf{g}, \mathbf{z} \rangle.$$

Using Lemma 2.9 and the equivalence of sub-gaussian properties, for instance in [11, Lemma 5.5], we have

$$(\mathbb{E}(\xi - w(K))_+^p)^{1/p} \leq C\sqrt{p}.$$

Above $(\xi - w(K))_+ := \max(\xi - w(K), 0)$. Applying Minkowski's inequality gives

$$(\mathbb{E} \xi^p)^{1/p} \leq (\mathbb{E}(\xi - w(K))_+^p)^{1/p} + (\mathbb{E} w(K)^p)^{1/p} \leq C\sqrt{p} + w(K).$$

Combine this with Equation (2.11) to give the moment bound

$$(2.12) \quad (\mathbb{E}(Z')^p)^{1/p} \leq C \cdot \frac{\kappa(\sqrt{p} + w(K))}{\sqrt{m}}.$$

In order to convert this into a tail bound, fix $\beta > 0$ and let $p \in [\beta^2, \beta^2 + 1)$ be β^2 rounded to the next highest integer. Further, let $t = e \cdot (\mathbb{E}(Z')^p)^{1/p}$. Then, by Markov's inequality we have

$$(2.13) \quad P(Z' \geq t) \leq \frac{\mathbb{E}(Z')^p}{t^p} \leq e^{-\beta^2} \quad \text{where} \quad t \leq C \cdot \frac{\kappa(\beta + w(K))}{\sqrt{m}}$$

To complete the proof of the proposition, apply Lemma 2.8: The moment bound (2.12) with $p = 1$ controls $\mathbb{E}(Z')$ and the tail bound (2.13) controls the right-hand side of Equation (2.10).

3. PROOF OF THEOREM 1.1

First observe that the proof of Proposition 2.2 is independent of $\theta(t)$ and in particular it holds for the sign function. The theorem then follows easily from an analogue of Proposition 2.1:

Proposition 3.1 (Expectation). *Consider $\mathbf{x}, \mathbf{x}' \in S^{n-1}$. If \mathbf{x} satisfies $\|\mathbf{x}\|_\infty \leq \lambda/(C \mathbb{E}|a|^3)$, then*

$$|\mathbb{E} f_{\mathbf{x}}(\mathbf{x}') - \langle \lambda \mathbf{x}, \mathbf{x}' \rangle| \leq \frac{C}{\lambda^{1/2}} \mathbb{E}|a|^3 \|\mathbf{x}\|_\infty^{1/2}.$$

To prove the proposition, we prove analogues of the lemmas used to prove the original proposition (Proposition 2.1).

Lemma 3.2. $|\langle \mathbf{v}_x, \mathbf{x} \rangle - \lambda| \leq C \mathbb{E}|a|^3 \|\mathbf{x}\|_3^3.$

Proof. Recall that by definition of \mathbf{x} ,

$$\langle \mathbf{v}_x, \mathbf{x} \rangle = \mathbb{E} \text{sign}(\langle \mathbf{a}, \mathbf{x} \rangle) \langle \mathbf{a}, \mathbf{x} \rangle = \mathbb{E} |\langle \mathbf{a}, \mathbf{x} \rangle|.$$

Note that $\lambda = \sqrt{2/\pi} = \mathbb{E}|g|$ and thus, to prove the lemma, we wish to bound the difference $|\mathbb{E} |\langle \mathbf{a}, \mathbf{x} \rangle| - \mathbb{E}|g||$. We have

$$|\mathbb{E} |\langle \mathbf{a}, \mathbf{x} \rangle| - \mathbb{E}|g|| = \left| \int_0^\infty P(|\langle \mathbf{a}, \mathbf{x} \rangle| \geq t) - P(|g| \geq t) dt \right| = 2 \left| \int_0^\infty P(\langle \mathbf{a}, \mathbf{x} \rangle \geq t) - P(g \geq t) dt \right|.$$

To conclude, we apply a Berry-Esseen result, for instance as in [9, Theorem 2.1.24], which bounds the above quantity by

$$C \sum_{i=1}^n \mathbb{E} |x_i a_i|^3 = C \mathbb{E} |a|^3 \|\mathbf{x}\|_3^3.$$

■

Observe that as a result of Lemma 3.2, Lemma 2.5 is proven as before but requiring only

$$\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_3^3 \leq \lambda/(2C \mathbb{E}|a|^3).$$

An analogue for Lemma 2.6, however, requires a different approach.

Lemma 3.3. $\|\mathbf{v}_x\|_\infty \leq C \mathbb{E}|a|^3 \|\mathbf{x}\|_\infty.$

Proof. Establishing the notation $\langle \mathbf{a}, \mathbf{x} \rangle = \sum_{k=1}^n a_k x_k$ where without loss of generality, $x_i \geq 0$, define for convenience $S = \sum_{k \neq i}^n a_k x_k$ and let F_S be the cumulative distribution function of S . Consider an arbitrary constant r .

$$\begin{aligned} |\mathbb{E} \theta(S + r x_i) \cdot r| &= \left| r \int_{\mathbb{R}} \text{sign}(t + r x_i) dF_S(t) \right| = |r| \left| \int_{t \geq -r x_i} dF_S(t) - \int_{t < -r x_i} dF_S(t) \right| \\ &= |r| |P(S \geq -r x_i) - P(S < -r x_i)| = |r| P(|S| \leq |r x_i|) \\ &\leq |r| P(|g| \leq |r x_i|) + |r| \cdot |P(|g| \leq |r x_i|) - P(|S| \leq |r x_i|)|. \end{aligned}$$

The second term in the last inequality may be bounded using the Berry-Esseen Theorem, which may be found for instance in [9, Theorem 2.1.30]. This gives

$$|\mathbb{E} \theta(S + rx_i) \cdot r| \leq |r| \left\{ \sqrt{\frac{2}{\pi}} |r| x_i + 2 \left(\sum_{k \neq i} x_k^2 \right)^{-3/2} \mathbb{E} |a|^3 \sum_{k \neq i} |x_k|^3 \right\}.$$

Note $\|\mathbf{x}\|_3^3 \leq \|\mathbf{x}\|_\infty \|\mathbf{x}\|_2^2 = \|\mathbf{x}\|_\infty \leq 1/8$, where the last inequality is by assumption. Then $x_i^3 \leq 1/8$, $x_i^2 \leq 1/4$, so that $\sum_{k \neq i} x_k^2 \geq 3/4$. Observing furthermore that $\|\mathbf{x}\|_\infty \geq \sum_{k \neq i} x_k^2 \|\mathbf{x}\|_\infty \geq \sum_{k \neq i} |x_k|^3$, we have the bound

$$|\mathbb{E} \theta(S + rx_i) \cdot r| \leq Cr^2 x_i + Cr \mathbb{E} |a|^3 \|\mathbf{x}\|_\infty.$$

We may express a single coordinate of $v_x = \mathbb{E} \theta(\langle \mathbf{a}, \mathbf{x} \rangle) \cdot \mathbf{a}$ as $\mathbb{E} \theta(\langle \mathbf{a}, \mathbf{x} \rangle) \cdot a_i$. Then,

$$\begin{aligned} |\mathbb{E} \theta(\langle \mathbf{a}, \mathbf{x} \rangle) \cdot a_i| &\leq \int_{\mathbb{R}} |\mathbb{E} \theta(S + tx_i) \cdot t| dF_{a_i}(t) \\ &\leq \int_{\mathbb{R}} (Ct^2 x_i + C|t| \mathbb{E} |a|^3 \|\mathbf{x}\|_\infty) dF_{a_i}(t) \\ &= Cx_i \mathbb{E} a_i^2 + C \mathbb{E} |a|^3 \|\mathbf{x}\|_\infty \mathbb{E} |a_i| \\ &\leq Cx_i + C \mathbb{E} |a|^3 \|\mathbf{x}\|_\infty. \end{aligned}$$

Observing that $\mathbb{E} |a|^3 \geq \mathbb{E} a^2 = 1$ completes the proof of the lemma. ■

Defining $\mathbf{z} = \mathbf{v}_x / \|\mathbf{v}_x\|_2$, applying Lemma 2.5 and Lemma 3.3 yields $\|\mathbf{z}\|_3^3 \leq \|\mathbf{z}\|_\infty \leq CE|a|^3 \|\mathbf{x}\|_\infty / \lambda$. Hence,

$$\begin{aligned} \|\mathbf{v}_x\|_2 &= \langle \mathbf{v}_x, \mathbf{v}_x / \|\mathbf{v}_x\|_2 \rangle = \mathbb{E} \text{sign}(\langle \mathbf{a}, \mathbf{x} \rangle) \langle \mathbf{a}, \mathbf{z} \rangle \\ &\leq \mathbb{E} \text{sign}(\langle \mathbf{a}, \mathbf{z} \rangle) \langle \mathbf{a}, \mathbf{z} \rangle = \langle \mathbf{v}_z, \mathbf{z} \rangle \\ &\leq \lambda + C \mathbb{E} |a|^3 \|\mathbf{z}\|_3^3 \leq \lambda + \frac{C}{\lambda} (\mathbb{E} |a|^3)^2 \|\mathbf{x}\|_\infty. \end{aligned}$$

Combining results, we have

$$\begin{aligned} |\langle \mathbf{v}_x, \mathbf{x}' \rangle - \langle \lambda \mathbf{x}, \mathbf{x}' \rangle|^2 &= \|\mathbf{v}_x\|^2 - \lambda^2 + 2\lambda(\lambda - \langle \mathbf{v}_x, \mathbf{x} \rangle) \\ &= (\|\mathbf{v}_x\| + \lambda)(\|\mathbf{v}_x\| - \lambda) + 2\lambda(\lambda - \langle \mathbf{v}_x, \mathbf{x} \rangle) \\ &\leq C \left(\frac{1}{\lambda} (\mathbb{E} |a|^3)^2 \|\mathbf{x}\|_\infty + \mathbb{E} |a|^3 \|\mathbf{x}\|_3^3 \right). \end{aligned}$$

Recalling that $\|\mathbf{x}\|_\infty \geq \|\mathbf{x}\|_3^3$, we may collect terms to conclude

$$|\mathbb{E} f_{\mathbf{x}}(\mathbf{x}') - \langle \lambda \mathbf{x}, \mathbf{x}' \rangle| \leq \frac{C}{\lambda^{1/2}} \mathbb{E} |a|^3 \|\mathbf{x}\|_\infty^{1/2}.$$

This completes the proof of Proposition 3.1.

Theorem 1.1 follows as in the proof of Theorem 1.3 in Section 2 and by noting $\lambda = \sqrt{2/\pi}$.

4. CONCLUSION

In contrast to standard compressed sensing, one-bit compressed sensing is infeasible when the measurement vectors are Bernoulli and the signal is extremely sparse. Nevertheless, we show that when the signal is sparse, but not overly sparse, it may be recovered from Bernoulli (or more generally, sub-gaussian) one-bit measurements. To our knowledge, these are the first theoretical results in one-bit compressed sensing that specifically allow non-Gaussian measurements.

APPENDIX

4.1. Proof of Lemma 2.3. We apply Lindeberg replacement argument in a way similar to [10, Proposition D.2]. Define $v_j = (x_j, z_j)$, and let $\mathbf{g} \in \mathbb{R}^n$ be a vector of independent standard Gaussian variables. Define $S_i = \sum_{j=1}^{i-1} a_j v_j + \sum_{j=i+1}^n g_j v_j$ and $\phi(v) = \theta(x)z$ (where $v = (x, z)$). Then note by telescoping,

$$\begin{aligned} |\mathbb{E} \theta(\langle \mathbf{a}, \mathbf{x} \rangle) \langle \mathbf{a}, \mathbf{z} \rangle - \mathbb{E} \theta(\langle \mathbf{g}, \mathbf{x} \rangle) \langle \mathbf{g}, \mathbf{z} \rangle| &= \left| \mathbb{E} \phi \left(\sum_{j=1}^n a_j v_j \right) - \mathbb{E} \phi \left(\sum_{j=1}^n g_j v_j \right) \right| \\ &\leq \sum_{i=1}^N |\mathbb{E} \phi(S_i + a_i v_i) - \mathbb{E} \phi(S_i + g_i v_i)|. \end{aligned}$$

By Taylor's theorem with remainder, we have

$$\phi(S_i + a_i v_i) = \phi(S_i) + \sum_{|\alpha|=1} (a_i v_i)^\alpha \partial^\alpha \phi(S_i) + \frac{1}{2} \sum_{|\alpha|=2} (a_i v_i)^\alpha \partial^\alpha \phi(S_i) + \frac{1}{6} \sum_{|\alpha|=3} (a_i v_i)^\alpha \partial^\alpha \phi(S'_i)$$

for some S'_i on the line segment joining S_i and $S_i + a_i v_i$. A similar result holds for $\phi(S_i + g_i v_i)$ with respective S''_i . Observe that since $\mathbb{E} a = \mathbb{E} g = 0$ and $\mathbb{E} a^2 = \mathbb{E} g^2 = 1$, the zeroth to second order terms cancel upon taking expectations in the difference

$$|\mathbb{E} \phi(S_i + a_i v_i) - \mathbb{E} \phi(S_i + g_i v_i)| = \frac{1}{6} \left| \mathbb{E} \sum_{|\alpha|=3} (a_i v_i)^\alpha \partial^\alpha \phi(S'_i) - \mathbb{E} \sum_{|\alpha|=3} (g_i v_i)^\alpha \partial^\alpha \phi(S''_i) \right|.$$

Consider the first expectation on the righthand side. Observe that the partials in the error vanish except when at most one partial is taken on the second argument of ϕ , yielding either $\theta''(x)$ or $\theta'''(x)z$. Furthermore, note that since S'_i is on the line segment joining S_i and $S_i + a_i v_i$, we may apply the bound $|(S'_i)_2| \leq |(S_i)_2| + |a_i z_i|$ to conclude

$$\begin{aligned} \mathbb{E} \left| \sum_{|\alpha|=3} (a_i v_i)^\alpha \partial^\alpha \phi(S'_i) \right| &\leq \mathbb{E} \sum_{|\alpha|=3} |a_i^3 v_i^\alpha| (\|\theta''\|_\infty + \|\theta'''\|_\infty (|(S_i)_2| + |a_i z_i|)) \\ &= (|x_i| + |z_i|)^3 (\tau_2 \mathbb{E} |a_i|^3 + \tau_3 (\mathbb{E} |(S_i)_2 a_i^3| + |z_i| \mathbb{E} a_i^4)) \end{aligned}$$

Observe that $(S_i)_2$ and a_i are independent, and $(\mathbb{E} |(S_i)_2|^2)^2 \leq \mathbb{E} (S_i)_2^2 \leq 1$ by Cauchy-Schwarz and that the variance of an independent sum is a sum of variances. Further observing that $|z_i| \leq 1$ and $\mathbb{E} |a|^3 \leq \mathbb{E} a^4$, we may collect terms to conclude

$$\mathbb{E} \left| \sum_{|\alpha|=3} (a_i v_i)^\alpha \partial^\alpha \phi(S'_i) \right| \leq (|x_i| + |z_i|)^3 (\tau_2 + 2\tau_3) \mathbb{E} a_i^4.$$

A similar bound follows for the remainder from the Gaussian expansion, so that summing over i from 1 to n , and observing that the Gaussian remainder can be absorbed since $\mathbb{E} a^4 \geq \mathbb{E} a^2 = 1$,

$$\left| \mathbb{E} \theta(\langle \mathbf{a}, \mathbf{x} \rangle) \langle \mathbf{a}, \mathbf{z} \rangle - \mathbb{E} \theta(\langle \mathbf{g}, \mathbf{x} \rangle) \langle \mathbf{g}, \mathbf{z} \rangle \right| \leq C(\tau_2 + \tau_3) \mathbb{E} a^4 \|\mathbf{x}\| + \|\mathbf{z}\|_3^3,$$

which completes the proof of the lemma.

4.2. Total variation: sign function. We consider the setting of Theorem 1.1, where $\theta(t) = \text{sign}(t)$, with the additional assumption that $\|a - g\|_{TV} \leq \varepsilon$.

Theorem 4.1 (Estimating a signal with no noise). *We remain in the setting of Theorem 1.1 with the additional condition $\|a - g\|_{TV} \leq \varepsilon$. Then for each $\beta > 0$, with probability at least $1 - 4e^{-\beta^2}$, the solution $\hat{\mathbf{x}}$ to the optimization problem (1.2) satisfies*

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 \leq C\sqrt{\kappa}\varepsilon^{1/8} + \frac{C\kappa}{\lambda\sqrt{m}}(w(K) + \beta).$$

As with Theorem 1.1, the main result follows easily from an analogue of Proposition 2.1. Below, $\lambda = \sqrt{2/\pi}$.

Proposition 4.2 (Expectation). *For $\mathbf{x}, \mathbf{x}' \in S^{n-1}$,*

$$|\mathbb{E} f_{\mathbf{x}}(\mathbf{x}') - \langle \lambda \mathbf{x}, \mathbf{x}' \rangle| \leq C(\mathbb{E} a^4)^{1/8} \varepsilon^{1/8}.$$

To prove the proposition, we prove an analogue of a lemma used to prove the Proposition 2.1.

Lemma 4.3. $|\langle \mathbf{v}_{\mathbf{x}}, \mathbf{x} \rangle - \lambda| = \left| \mathbb{E} |\langle \mathbf{a}, \mathbf{x} \rangle| - \sqrt{\frac{2}{\pi}} \right| \leq 4(\mathbb{E} a^4 + \mathbb{E} g^4)^{1/4} \varepsilon^{1/4}.$

Proof. We first prove a variant of the Berry-Esseen result on expectations, applying Lindeberg replacement. Define $S_i = \sum_{j=1}^{i-1} a_j x_j + \sum_{j=i+1}^n g_j x_j$, and $\phi(x)$ to be a twice differentiable function. We will later use an approximation argument to replace ϕ by the absolute value function.

Note by telescoping,

$$\begin{aligned} |\mathbb{E} \phi(\langle \mathbf{a}, \mathbf{x} \rangle) - \mathbb{E} \phi(\langle \mathbf{g}, \mathbf{x} \rangle)| &= \left| \mathbb{E} \phi\left(\sum_{i=1}^N a_i x_i\right) - \mathbb{E} \phi\left(\sum_{i=1}^N g_i x_i\right) \right| \\ &\leq \sum_{i=1}^N |\mathbb{E} \phi(S_i + a_i x_i) - \mathbb{E} \phi(S_i + g_i x_i)|. \end{aligned}$$

For convenience, dropping subscripts, we now wish to bound $|\mathbb{E} \phi(S + ax) - \mathbb{E} \phi(S + gx)|$.

By Taylor's theorem with remainder, we have

$$\phi(S + ax) = \phi(S) + ax\phi'(S) + R(S, ax)$$

where $|R(S, ax)| \leq (ax)^2 \|\phi''\|_{\infty}/2$. A similar result holds for $\phi(S + gx)$.

Split $R(S, x)$ into $R_+(S, x) \geq 0$ and $R_-(S, x) \leq 0$. Observe that since $\mathbb{E} a = \mathbb{E} g = 0$, the zeroth and first order terms cancel upon taking expectations in the difference

$$\begin{aligned} |\mathbb{E} \phi(S + ax) - \mathbb{E} \phi(S + gx)| &= |\mathbb{E} R(S, ax) - \mathbb{E} R(S, gx)| \\ &\leq |\mathbb{E} R_+(S, ax) - \mathbb{E} R_+(S, gx)| + |\mathbb{E} R_-(S, ax) - \mathbb{E} R_-(S, gx)|. \end{aligned}$$

Consider the difference with R_+ . We will apply the assumption $\|a - g\|_{TV} \leq \varepsilon$. First, observe that S is independent of both a and g and may be viewed as a constant. Viewing for instance $R_+(S, ax)$ as a function of a ,

$$\left| \int_0^M P(R_+(S, ax) > t) dt - \int_0^M P(R_+(S, gx) > t) dt \right| \leq M\varepsilon.$$

Then, consider the tail of the first integral:

$$\begin{aligned} \int_M^\infty P(R_+(S, ax) > t) dt &\leq \int_M^\infty \frac{\mathbb{E}(R_+(S, ax)^2)}{t^2} dt \\ &= \frac{\mathbb{E} R_+(S, ax)^2}{M} \leq \frac{x^4 \mathbb{E} a^4 \|\phi''\|_\infty^2}{4M}. \end{aligned}$$

The Gaussian tail yields a similar error. Hence, optimizing over M by choosing

$$M = \frac{x^2(\mathbb{E} a^4 + \mathbb{E} g^4)^{1/2} \|\phi''\|_\infty}{2\sqrt{\varepsilon}}$$

we have overall error

$$|\mathbb{E} R_+(S, ax) - \mathbb{E} R_+(S, gx)| \leq x^2(\mathbb{E} a^4 + \mathbb{E} g^4)^{1/2} \|\phi''\|_\infty \sqrt{\varepsilon}.$$

The same holds for the difference with R_- . Finally, summing over the n indices, and using that $\|x\|_2 = 1$,

$$|\mathbb{E} \phi(\langle \mathbf{a}, \mathbf{x} \rangle) - \mathbb{E} \phi(\langle \mathbf{g}, \mathbf{x} \rangle)| \leq 2(\mathbb{E} a^4 + \mathbb{E} g^4)^{1/2} \|\phi''\|_\infty \sqrt{\varepsilon}.$$

Second, we approximate the absolute value using $\phi(x) := \sqrt{c + x^2} \approx |x|$. Observe for instance that $|\mathbb{E} |\langle a, x \rangle| - \mathbb{E} \phi(\langle a, x \rangle)| \leq \sqrt{c}$, and likewise with g in the place of a . Evaluating $\phi''(x) = c/(c + x^2)^{3/2}$ with a maximum of $1/\sqrt{c}$ at $x = 0$, we may conclude

$$|\langle \mathbf{v}_x, \mathbf{x} \rangle - \lambda| = \left| \mathbb{E} |\langle \mathbf{a}, \mathbf{x} \rangle| - \mathbb{E} |\langle \mathbf{g}, \mathbf{x} \rangle| \right| \leq 2\sqrt{c} + 2(\mathbb{E} a^4 + \mathbb{E} g^4)^{1/2} \sqrt{\frac{\varepsilon}{c}}.$$

Choosing $\sqrt{c} = (\mathbb{E} a^4 + \mathbb{E} g^4)^{1/4} \varepsilon^{1/4}$ completes the proof of the lemma. ■

We now proceed to bound $\|\mathbf{v}_x\|_2$, thus obtaining the second geometric constraint required in the proof of the proposition. We apply lemma 4.3 with $\mathbf{v}_x/\|\mathbf{v}_x\|_2$ in the place of \mathbf{x} :

$$\begin{aligned} \|\mathbf{v}_x\|_2 &= \langle \mathbf{v}_x, \mathbf{v}_x/\|\mathbf{v}_x\|_2 \rangle = \mathbb{E} \text{sign}(\langle \mathbf{a}, \mathbf{x} \rangle) \langle \mathbf{a}, \mathbf{z} \rangle \\ &\leq \mathbb{E} \text{sign}(\langle \mathbf{a}, \mathbf{z} \rangle) \langle \mathbf{a}, \mathbf{z} \rangle = \langle \mathbf{v}_z, \mathbf{z} \rangle \leq \lambda + 4(\mathbb{E} a^4 + \mathbb{E} g^4)^{1/4} \varepsilon^{1/4}. \end{aligned}$$

An additional direct application of the lemma yields

$$\begin{aligned} |\langle \mathbf{v}_x, \mathbf{x}' \rangle - \langle \lambda \mathbf{x}, \mathbf{x}' \rangle|^2 &= \|\mathbf{v}_x\|_2^2 - \lambda^2 + 2\lambda(\lambda - \langle \mathbf{v}_x, \mathbf{x} \rangle) \\ &= (\|\mathbf{v}_x\|_2 + \lambda)(\|\mathbf{v}_x\|_2 - \lambda) + 2\lambda(\lambda - \langle \mathbf{v}_x, \mathbf{x} \rangle) \\ &\leq 16(\mathbb{E} a^4 + \mathbb{E} g^4)^{1/4} \varepsilon^{1/4}. \end{aligned}$$

Proposition 4.2 is a consequence of absorbing constants, and Theorem 4.1 follows as in the proof of Theorem 1.3 in Section 2.

4.3. Total variation: smooth noise model. We consider the setting of Theorem 1.3, with the additional assumption that $\|a - g\|_{TV} \leq \varepsilon$. We also relax the assumption on $\theta(t)$, defined as in (1.6), to $\theta(t) \in C^2$.

Theorem 4.4 (Estimating a signal with noise). *We remain in the setting of Theorem 1.3 with the additional condition $\|a - g\|_{TV} \leq \varepsilon$, and also relax the condition on $\theta(t)$ to $\theta(t) \in C^2$. Then for each $\beta > 0$, with probability at least $1 - 4e^{-\beta^2}$, the solution $\hat{\mathbf{x}}$ to the optimization problem (1.2) satisfies*

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 \leq C \left((\kappa^3 + 1)(\tau_1 + \tau_2)\sqrt{\varepsilon} + \frac{\kappa}{\lambda\sqrt{m}}(w(K) + \beta) \right).$$

As with Theorem 1.1, the main result follows easily from an analogue of Proposition 2.1:

Proposition 4.5 (Expectation). *For $\mathbf{x}, \mathbf{x}' \in S^{n-1}$,*

$$|\mathbb{E} f_{\mathbf{x}}(\mathbf{x}') - \langle \lambda \mathbf{x}, \mathbf{x}' \rangle| \leq 8(\mathbb{E} a^6 + \mathbb{E} g^6)^{1/2}(\tau_1 + \tau_2)\sqrt{\varepsilon}.$$

Because we intend to have no dependence on \mathbf{x} , the required generality of \mathbf{x}' is no additional burden. As a result, it is possible to prove the proposition directly.

Proof. Recalling [2, Lemma 4.1], observe that the left hand side of the inequality is expressible as

$$|\mathbb{E} f_{\mathbf{x}}(\mathbf{x}') - \langle \lambda \mathbf{x}, \mathbf{x}' \rangle| = |\mathbb{E} \theta(\langle \mathbf{a}, \mathbf{x} \rangle) \langle \mathbf{a}, \mathbf{x}' \rangle - \mathbb{E} \theta(\langle \mathbf{g}, \mathbf{x} \rangle) \langle \mathbf{g}, \mathbf{x}' \rangle|.$$

The statement of the proposition becomes similar to that of Lemma 2.3. Using the same notation and proceeding as in its proof (hence for instance using \mathbf{z} in place of $v\mathbf{x}'$), we apply Lindeberg replacement:

$$|\mathbb{E} \theta(\langle \mathbf{a}, \mathbf{x} \rangle) \langle \mathbf{a}, \mathbf{z} \rangle - \mathbb{E} \theta(\langle \mathbf{g}, \mathbf{x} \rangle) \langle \mathbf{g}, \mathbf{z} \rangle| \leq \sum_{i=1}^N |\mathbb{E} \phi(S_i + a_i v_i) - \mathbb{E} \phi(S_i + g_i v_i)|.$$

As before, we Taylor expand, except only to second order error:

$$\phi(S_i + a_i v_i) = \phi(S_i) + \sum_{|\alpha|=1} (a_i v_i)^\alpha \partial^\alpha \phi(S_i) + R(S_i, a_i v_i)$$

where $R(S_i, a_i v_i) = \frac{1}{2} \sum_{|\alpha|=2} (a_i v_i)^\alpha \partial^\alpha \phi(S'_i)$ for some S'_i on the line segment joining S_i and $S_i + a_i v_i$. A similar result holds with $\phi(S_i + g_i v_i)$, with respective S''_i .

Split $R(S, v)$ into $R_+(S, v) \geq 0$ and $R_-(S, v) \geq 0$. Observe that since $\mathbb{E} a = \mathbb{E} g = 0$, the zeroth and first order terms cancel upon taking expectations in the difference

$$\begin{aligned} & |\mathbb{E} \phi(S_i + a_i v_i) - \mathbb{E} \phi(S_i + g_i v_i)| = |\mathbb{E} R(S_i, a_i v_i) - \mathbb{E} R(S_i, g_i v_i)| \\ & \leq |\mathbb{E} R_+(S_i, a_i v_i) - \mathbb{E} R_+(S_i, g_i v_i)| + |\mathbb{E} R_-(S_i, a_i v_i) - \mathbb{E} R_-(S_i, g_i v_i)|. \end{aligned}$$

Consider the difference containing R_+ . We will apply the assumption $\|a - g\|_{TV} \leq \varepsilon$. First, observe that S_i is independent of both a_i and g_i and may be viewed as a constant (by conditioning

on it). Viewing for instance $R_+(S_i, a_i v_i)$ as a function of a_i ,

$$\left| \int_0^M P(R_+(S_i, a_i v_i) > t) dt - \int_0^M P(R_+(S_i, g_i v_i) > t) dt \right| \leq M\varepsilon.$$

Then, consider the tail of the first integral:

$$\int_M^\infty P(R_+(S_i, a_i v_i) > t) dt \leq \int_M^\infty \frac{\mathbb{E}(R_+(S_i, a_i v_i)^2)}{t^2} dt = \frac{\mathbb{E} R_+(S_i, a_i v_i)^2}{M}.$$

Recall the explicit form of the remainder and observe that the partials in the error vanish except when at most one partial is taken on the second argument of ϕ , yielding either $\theta'(x)$ or $\theta''(x)z$. Furthermore, note that since S'_i is on the line segment joining S_i and $S_i + a_i v_i$, we may apply the bound $|(S'_i)_2| \leq |(S_i)_2| + |a_i z_i|$ to conclude

$$\begin{aligned} \mathbb{E} 4R_+(S_i, a_i v_i)^2 &\leq \mathbb{E} 4R(S_i, a_i v_i)^2 \leq \mathbb{E} \left(\sum_{|\alpha|=2} a_i^2 |v_i^\alpha| (\|\theta'\|_\infty + \|\theta''\|_\infty (|(S_i)_2| + |a_i z_i|)) \right)^2 \\ &= \mathbb{E} (a_i^2 (|x_i| + |z_i|)^2 (\tau_1 + \tau_2 (|(S_i)_2| + |z_i a_i|)))^2 \end{aligned}$$

Observe that $(S_i)_2$ and a_i are independent, and $(\mathbb{E} |(S_i)_2|)^2 \leq \mathbb{E}(S_i)_2^2 \leq 1$ by Cauchy-Schwarz and that the variance of an independent sum is a sum of variances. Further observing that $|z_i| \leq 1$ and for instance $\mathbb{E} |a|^5 \leq \mathbb{E} a^6$, rearranging and collecting terms yields

$$\begin{aligned} \mathbb{E} 4R_+(S_i, a_i v_i)^2 &\leq (|x_i| + |z_i|)^4 (4\tau_2^2 \mathbb{E} a_i^6 + \tau_1^2 \mathbb{E} a_i^4 + 4\tau_1 \tau_2 \mathbb{E} |a_i|^5) \\ &\leq 4(|x_i| + |z_i|)^4 (\tau_1 + \tau_2)^2 \mathbb{E} a_i^6. \end{aligned}$$

The Gaussian tail yields a similar error. Hence, optimizing over M by choosing

$$M = \frac{1}{\sqrt{\varepsilon}} (|x_i| + |z_i|)^2 (\mathbb{E} a^6 + \mathbb{E} g^6)^{1/2} (\tau_1 + \tau_2)$$

we have overall error

$$|\mathbb{E} R_+(S_i, a_i v_i) - \mathbb{E} R_+(S_i, g_i v_i)| \leq 2(|x_i| + |z_i|)^2 (\mathbb{E} a^6 + \mathbb{E} g^6)^{1/2} (\tau_1 + \tau_2) \sqrt{\varepsilon}.$$

The same holds for the difference with R_- . Finally, summing over the n indices, and using that $\|\mathbf{x}\|_2 = 1$ and $\|\mathbf{z}\|_2 = 1$,

$$|\mathbb{E} \phi(\langle \mathbf{a}, \mathbf{x} \rangle) - \mathbb{E} \phi(\langle \mathbf{g}, \mathbf{x} \rangle)| \leq 8(\mathbb{E} a^6 + \mathbb{E} g^6)^{1/2} (\tau_1 + \tau_2) \sqrt{\varepsilon},$$

which concludes the proof of the proposition. ■

We can further simplify the error expression in Proposition 4.5 by observing that

$$(\mathbb{E} a^6 + \mathbb{E} g^6)^{1/2} \leq C(\kappa^3 + 1).$$

Then Theorem 4.4 follows as in the proof of Theorem 1.3 in Section 2.

REFERENCES

- [1] BOUFONOS, P. T., AND BARANIUK, R. G. 1-Bit compressive sensing. In *42nd Annual Conference on Information Sciences and Systems (CISS)* (Mar. 2008).
- [2] ELDAR, C., AND KUTYNIOK, G., Eds. *Compressed Sensing: Theory and applications*. Cambridge University Press, 2012.
- [3] GUPTA, A., NOWAK, R., AND RECHT, B. Sample complexity for 1-bit compressed sensing and sparse classification. In *International Symposium on Information Theory (ISIT)* (2010), IEEE.
- [4] JACQUES, L., LASKA, J. N., BOUFONOS, P. T., AND BARANIUK, R. G. Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors. Preprint. Available at <http://arxiv.org/abs/1104.3160>.
- [5] LEDOUX, M. *The concentration of measure phenomenon*. American Mathematical Society, Providence, 2001.
- [6] LEDOUX, M., AND TALAGRAND, M. *Probability in Banach Spaces: isoperimetry and processes*. Springer-Verlag, Berlin, 1991.
- [7] PLAN, Y., AND VERSHYNIN, R. One-bit compressed sensing by linear programming. Preprint. Available at <http://arxiv.org/abs/1109.4299>.
- [8] PLAN, Y., AND VERSHYNIN, R. Robust 1-bit compressed sensing and sparse logistic regression: A convex programming approach. Preprint. Available at <http://arxiv.org/abs/1202.1212>.
- [9] STROOK, D. *Probability Theory: an analytic view*. Cambridge University Press, 1993.
- [10] TAO, T., AND VU, V. Random matrices: The distribution of the smallest singular values. *Geometric And Functional Analysis* 20, 1 (2010), 260–297.
- [11] VERSHYNIN, R. Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing: Theory and Applications*, Y. Eldar and G. Kutyniok, Eds. Cambridge University Press, 2012.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MICHIGAN, 530 CHURCH ST., ANN ARBOR, MI 48109, U.S.A.

E-mail address: {aflapan, yplan, romanv}@umich.edu

E-mail address: aai@princeton.edu